UTTERANCE AND OBJECTIVE: ISSUES IN NATURAL LANGUAGE COMMUNICATION

Technical Note 188R

June 1979

By: Barbara J. Grosz, Program Director

Artificial Intelligence Center
Computer Science and Technology Division

This paper appears in the *AI Magazine*, Volume I, No. 1 (Spring 1980). An earlier version appears in *Proceedings of the Sixth International Joint Conference on Artificial Intelligence*, Tokyo, Japan (August 20-23, 1979).

Preparation of this paper was supported by the National Science Foundation under Grant No. MCS76-220004, and the Defense Advanced Research Projects Agency under Contract N00039-79-C0118 with the Naval Electronic Systems Command to SRI International.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 1979		**00-06-1979 to 00-06-1979**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Utterance and Objective: Issues in Natural Language Communication	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SRI International,333 Ravenswood Avenue,Menlo Park,CA,94025**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**11**	

UTTERANCE AND OBJECTIVE:

Issues in Natural Language Communication

Barbara J. Grosz
Artificial Intelligence Center
SRI International
Menlo Park, CA 94025

I. Introduction

Two premises, reflected in the title, underlie the perspective from which I will consider research in natural language processing in this paper.* First, progress on building computer systems that process natural languages in any meaningful sense (i.e., systems that interact reasonably with people in natural language) requires considering language as part of a larger communicative situation. In this larger situation, the participants in a conversation and their states of mind are as important to the interpretation of an utterance as the linguistic expressions from which it is formed. A central concern when language is considered as communication is its function in building and using shared models of the world. Indeed, the notion of a shared model is inherent in the word "communicate," which is derived from the Latin *communicare*, "to make common."

Second, as the phrase "utterance and objective" suggests, regarding language as communication requires consideration of what is said literally, what is intended, and the relationship between the two. Recently, the emphasis in research in natural language processing has begun to shift from an analysis of utterances as isolated linguistic phenomena to a consideration of how people use utterances to achieve certain objectives. But, in considering objectives, it is important not to ignore the utterances themselves. A consideration of a speaker's underlying goals and motivations is critical, but so is an analysis of the particular way in which that speaker expresses his thoughts. (I will use "speaker" and "hearer" to refer respectively to the producer of an utterance and the interpreter of that utterance. Although the particular communicative environment constrains the set of linguistic and nonlinguistic devices a speaker may use (Rubin, 1977), I will ignore the differences and concentrate on those problems that are common across environments.) The choice of expression has implications for such things as what other entities may be discussed in the ensuing discourse, what the speaker's underlying beliefs (including his beliefs about the hearer) are, and what social relationship the speaker and hearer have. The reason for conjoining "utterance" and

---

*This is a revision of a paper presented at the Sixth International Conference on Artificial Intelligence, Tokyo, Japan, August 20-24, 1979. Preparation of this paper was supported by the National Science Foundation under Grant No. MCS76-220004, and the Defense Advanced Research Projects Agency under Contract N00039-79-C0118 with the Naval Electronic Systems Command.

"objective" in the title of this paper is to emphasize the importance of considering both. (The similarity to *Word and Object* (Quine, 1960) is not entirely accidental. It is intended to highlight a major shift in the context in which questions about language and meaning should be considered. I believe the issues Quine raised can be addressed effectively only in this larger context.)

In the remainder of this paper I will examine three consequences of these claims for the development of language processing theories and the construction of language processing systems.

- Language processing requires a combination of language-specific mechanisms and general commonsense reasoning mechanisms. Specifying these mechanisms and their interactions constitutes a major research area.
- Because discourse involves multiple separate agents with differing conceptions of the world, language systems must be able to represent the beliefs and knowledge of multiple individual agents. The reasoning procedures that operate on these representations must be able to handle such separate beliefs. Furthermore, they must be able to operate on incomplete and sometimes inconsistent information.
- Utterances are multifaceted; they must be viewed as having effects along multiple dimensions. As a result, commonsense reasoning (especially planning) procedures must be able to handle situations that involve actions having multiple effects.

II. Monkeys, Bananas, and Communication

To illustrate some of the current problems in natural language processing, I will consider a variant of the "monkey and bananas" problem (McCarthy, 1968), the original version of which is substantially as follows: There is a monkey in a room in which a bunch of bananas is hanging from the ceiling, out of reach of the monkey. There is also a box in one corner of the room. The monkey's problem is to figure out what sequence of actions will get him the bananas. For a while at least, this problem was a favorite test case for automatic problem solvers, and there are several descriptions of how it can be solved by machine (e.g., see Nilsson, 1971). The variation I will discuss introduces a second monkey, the need for some communication to take place, and a change of scene to a tropical forest containing banana trees. To begin, I leave unspecified the relationship between the two monkeys and consider the short segment of hypothetical dialogue in Illustration 1:

(1) monkey1: I'm hungry.
(2) monkey2: There's a stick under the old rubber tree.

If monkey1 interprets monkey2's response as most current Artificial Intelligence (AI) natural language processing systems would, he might respond with something like, "I can't eat a

Illustration 1.

stick" or "I know, so what?" and, unless monkey2 helped him out, monkey1 would go hungry. Although there are a few systems now that might, with suitable tweaking, be able to get far enough for a response that indicates they have figured out that monkey2 intends for the stick to be used to knock down the bananas, there are no systems yet that would be able to understand most of the nuances of this response. For example, it implies not only that monkey2 has a plan for using the stick, but also that he expects monkey1 either to have a similar plan or to be able to figure one out once he has been told about the stick.

There is a corresponding amount of sophisticated knowledge and reasoning involved in monkey1's recognition of this request. To interpret "I'm hungry" correctly, monkey2 must recognize that a declarative statement is being used to issue a request. The robot's response in the dialogue of Illustration 2 reflects a lack of such recognition. It is inappropriate because it addresses the literal content of the monkey's statement rather than considering why he uttered it. (Notice that such a response might be appropriate in a different situation. For example, if the monkey were already eating a banana, "I'm hungry" could serve to explain why he was eating and "I understand" might serve as an acceptance of this explanation.)

A similar problem can arise with more explicit requests, like that given by the monkey in Illustration 3. Although the fact that the monkey is making a request is explicit here, the intent of his request must still be inferred. "Can you help me.." is an indirect request for assistance, not a question about the robot's capabilities. Again the response is inappropriate because it addresses the literal content of the message rather than the intent that underlies it. Taking queries literally is a major cause of inappropriate responses by natural language processing systems (and computer systems more generally).

If we complicate the scenario just slightly, we can create a situation that would cause trouble for all current natural language processing systems. In particular, suppose that the tree the stick is under is not a rubber tree, but rather a different sort of tree. Monkey2 might still use the phrase "the rubber tree," either by mistake or design, if he believes the phrase will suffice to enable monkey1 to identify the tree (c.f. Donnellan, 1977). No current AI natural language processing system

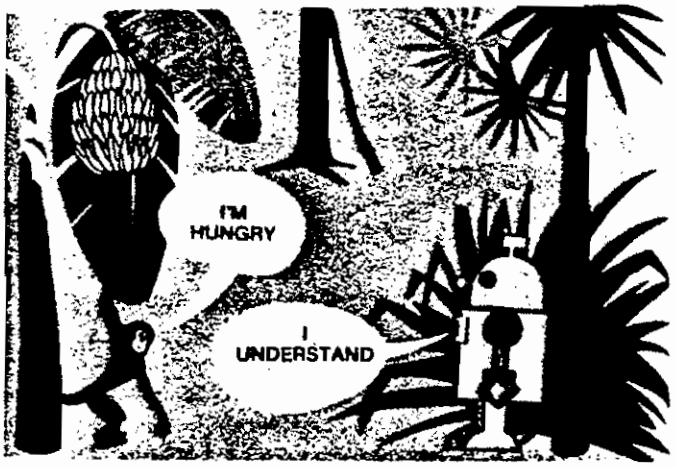

Illustration 2.

Illustration 3.

would be able to figure out where the stick is. Their responses, at best, would be like monkey1 saying, "Whaddayamean? There aren't any rubber trees in this forest." But referring expressions that do not accurately describe the entities they are intended to identify are typical of the sort of thing that occurs all the time in conversations between humans. The question is what it will take to get computer systems closer to being able to handle these sorts of phenomena.

In the remainder of this paper I will examine some of the research issues that need to be addressed to bring us closer to understanding why talking monkeys don't go hungry. Many of the problems that must be confronted are not confined solely to natural language processing but fall under the larger purview of AI more generally. Many critical language processing issues arise from our limited knowledge of how common-sense reasoning—which includes deduction, plausible reasoning, planning, and plan recognition—can be captured in a computational system. Consequently, research in natural language processing and research in common-sense reasoning must be tightly coordinated in the next few years.

A major source of the inadequacies of current common-sense reasoning mechanisms, when considered as possible components in a natural language processing system, is the following discrepancy. Research in problem solving and deduction has focused almost exclusively on problems that a single agent could solve alone. The need for communication arises with those problems that require the resources of multiple agents, problems that a single agent has insufficient power to solve alone. As a result, language processing is typically an issue in just those contexts where the aid of another agent is essential. To obtain that aid, the first agent must take into account the knowledge, capabilities, and goals of the second. In exchange for not needing quite as much knowledge or capability in the problem domain, the agent must have additional communication capabilities. For such problems, the option of proceeding without considering the independence of other agents and the need to communicate with them is not feasible. I believe this option is becoming less feasible as well for problem solving and deduction components used for other purposes within AI. Situations in which multiple robots must cooperate introduce similar complexity even if the communication itself can be carried out in a formal language. Sacerdoti (1978) discusses the usefulness of research in natural language processing for the construction of distributed artificial intelligence systems. The issues being raised in this paper are central AI issues; they provide evidence of the interconnectedness of natural language processing research and other research in AI.

III. The Processes of Interpretation

To illustrate how language-specific processes combine with general cognitive processes (i.e., common-sense reasoning) in the interpretation of an utterance, let us consider the first monkeys and bananas example in more detail. In the following analysis, a consideration of the states of mind of the speaker and hearer will play a critical role. Each participant in a conversation brings with him a cognitive state that includes such things as a focus of attention, a set of goals to be achieved or maintained and plans for achieving them, knowledge about the domain of discourse, knowledge about how language is used, and beliefs about the cognitive states of other agents, including other participants in the current conversation. An utterance conveys information about the speaker's state; its most immediate effect is to change the hearer's state.

It is useful to view natural language interpretation as being divided into two major interacting levels. On the first, the *linguistic analysis level*, the form of an utterance is analyzed to determine its context-independent attributes. Processes at this level are concerned with determining what information is contained in the utterance itself. On the second, the *assimilation level*, common-sense reasoning processes operating in the context of the current cognitive state of the hearer use these attributes to update the cognitive state and to determine what response to the utterance is required, if any. It is important to understand that the purpose of this separation is to elucidate the kinds of processes involved in interpretation. The actual flow of processing during interpretation entails a great deal of interaction among the processes in the different levels, and there are major research issues concerned with their coordination (e.g., see Robinson, 1980a; Walker, 1978).

To illustrate these levels, let us return to the example and consider the interpretation of monkey2's response (2),

"There's a stick under the old rubber tree," to monkey1's indirect request (1).

A. Linguistic Analysis

At this level, the parsing process that assigns syntactic structure to the utterance also assigns attributes to the various syntactic subphrases in the utterance and to the utterance as a whole. Many of these attributes are of a semantic or pragmatic nature. For example, the attributes of the phrase "the old rubber tree" might include

- The phrase is of syntactic class NP (noun phrase).
- The phrase is definitely determined.
- The phrase describes a t such that TREE(t) and OLD(t,T), where OLD and TREE are predicate symbols and the second argument to the predicate OLD indicates the set with respect to which age is evaluated.

I have left open the question of what happens with the modifier "rubber"; suffice it to say, the question of how it modifies cannot be resolved solely at the linguistic level. In general, the question of how much semantic specificity should be imposed at the linguistic level is an open research question.

Attributes of a complete utterance include such properties as its syntactic structure and the presuppositions (or, implicit assumptions) and assertions it makes. (Although what an utterance presupposes and asserts are not necessarily components of the intended meaning, the recognition of presuppositions and assertions is prerequisite to the assimilation level of processing.) Attributes of utterance (2) as a whole include:

- The utterance presupposes that there exists a t such that OLD(t,T) and TREE(t), and that the description "OLD(t,T)&TREE(t)" should allow t to be determined uniquely in the current context.
- The utterance asserts that there exists an s such that STICK(s).
- The utterance asserts that UNDER(s,t).

B. Assimilation

As attributes are extracted through the parsing process at the linguistic analysis level, common-sense reasoning processes begin to act on those attributes at the assimilation level. Two major activities are involved: completing the literal interpretation of an utterance in context, and drawing implications from that interpretation to discover the intended meaning.

For the example utterance (2), completing the literal interpretation in context involves the identification of the referent of the definite noun phrase, "the old rubber tree." The first attribute above indicates that a unique tree should be easily identified in context. Those objects currently in monkey1's focus of attention are examined (perhaps requiring sophisticated common-sense reasoning) to determine whether there is such a tree among them. Assume that none is found. It may be that only two kinds of trees are present in this forest, and that one kind, say gumgum trees, resemble rubber trees, and that of all the trees near the two monkeys only one is a gumgum tree. Monkey1 may tentatively assume that "rubber tree" matches "gumgum tree" closely enough to serve to identify this tree.

The sentence says there's a stick under the tree, so monkey1 might look under the tree and discover that, indeed, there is exactly one stick there. That stick must be the stick whose existence monkey2 was informing him of. The literal interpretation of the utterance is seen to be that the newly found stick is under the gumgum tree. (For more complex utterances, the process of completing the literal interpretation can involve determining the scopes of quantifiers and resolving various types of ambiguities.)

Knowing that the sentence presupposed the existence of a rubber tree and asserted the existence of a stick, monkey1 may infer that monkey2 believes these presuppositions. Thus, monkey1 comes to believe several new things about monkey2's beliefs; in particular, that he believes these two entities exist, and that he thinks the gumgum tree is a rubber tree, or at least thinks that this description can be used to identify the tree. This fact may be important in further communications. Monkey1 may also infer that because monkey2 has just mentioned the stick and the tree, they are in his focus of attention and that he (monkey2), too, should pay special attention to these objects. The stick may be of particular importance because it was the subject of a there-insertion sentence (a syntactic position of prominence) and has been newly introduced into his focus of attention.

The second major process of assimilation is to use common-sense reasoning to determine how the utterance fits into the current set of plans and goals. In general, this is a highly complex process.* For the particular example of interpreting utterance (2) in the context implied by utterance (1), monkey1 must determine what, "There's a stick under the rubber tree," has to do with his problem of getting something to eat. Briefly, he must see that the sentence emphasizes the stick and must know (or infer) that such sticks are often useful tools for getting things out of trees. He must infer that monkey2 intends for him to use this stick in conjunction with a standard plan for knocking down things to acquire some bananas and accomplish his (implicitly stated) goal of not being hungry.

IV. The Multifaceted Nature of Utterances

Just as an agent may perform physical actions intended to alter the physical state of his environment, he may perform linguistic actions (utter sentences) intended to alter the cognitive state of the hearer. To determine what objective an

*The complexity is well illustrated by the analysis of a set of therapeutic interviews in Labov and Fanshel (1977).

utterance is intended to achieve requires determining where that utterance fits in the speaker's plans. But because a single utterance may be used to achieve multiple effects simultaneously, the problem is more complex than either the analogy with physical actions or the preceding examples at first seem to suggest. (Physical actions may also have effects along multiple dimensions although they are not usually thought of as doing so. For example, the action of slamming a door in someone's face not only results in the door being closed, but also communicates anger.)

The discussion so far has concentrated on a single dimension of effect: the use of an utterance to achieve what I will call a domain goal, that is, to convey information about the domain of discourse. In this section I want to discuss two other dimensions along which an utterance can have effects—the social and the discourse—and look at some of the problems in interpretation and generation that arise from the multifaceted nature of utterances.*

To illustrate the three dimensions, consider the following utterance made by the hungry monkey in our illustrations (in this instance assume he sees the stick and realizes it can be used to knock down some bananas),

"Please hand me the stick."

At the domain level, the utterance expresses a proposition that might be written as HAND (MONKEY2, MONKEY1, S1), where MONKEY1 refers to monkey1 (the hungry monkey), MONKEY2 to monkey2, S1 to the stick under the tree, and HAND to the operation of transferring some object (given in the third argument) from one agent (first argument) to another (second argument) by hand. General domain information such as the taxonomic relationship that HAND is a kind of GIVE, and plan-based information about using the stick are an implicit part of the interpretation of the utterance along this dimension. At the social level, the utterance is a request; its imperative mood is modified by "please." At the discourse level, the utterance identifies and focuses on the stick S1.

The social dimension includes those aspects of an utterance that concern the establishment and maintenance of interpersonal relationships. This dimension of utterance (1), "I'm hungry," is easily seen when it is compared with such choices as

(3) "How can I get some of those blasted bananas?"
(4) "Can you help me get some bananas down?"
(5) "Get me a banana."

Each of these achieves the same domain goal, informing monkey2 of monkey1's desire to obtain some bananas, but utterance (1) does not convey the same familiarity as utterance (3) or the same level of frustration. (The bananas, after all, are not "blasted.") Similarly, utterance (4) makes the same request as utterance (5) but does so indirectly. A big monkey might use (5) to a small monkey and get a banana, but if a small monkey uttered it to a big monkey, he would more likely get a response like, "Not until you show some respect for your elders." A typical use of indirect speech acts like (4) is to moderate requests.

The social dimension is present in every discourse* and prevails in some (e.g., Hobbs, 1979). It has been largely ignored in natural language processing research to date. However, any analysis that translates the utterances in (1) and (3)-(5) only into requests for help getting food, misses a significant part of the meaning of each of the utterances. An assumption has been that some sort of neutral stance is possible. But, even the choice of the unmarked (neutral) case is a choice; not choosing is choosing not to choose (cf. Goffman, 1978). Although there are some serious philosophical issues raised by this dimension of utterances when considering communication between people and computers, I do not think we can continue to ignore it.

The *discourse dimension* includes those aspects of an utterance that derive from its participation in a coherent discourse—how the utterance relates to the utterances that preceded it and to what will follow. Although language is linear (only one word can be uttered at a time), the information a speaker has to convey typically is highly interconnected. As a result, the speaker must use multiple utterances to convey it. Each individual utterance must contain information that provides links to what went before and properly set the stage for what follows. Utterances that convey the same propositional content may differ widely in such things as the entities they indicate a speaker is focused on and hence may refer to later. As an extreme example, note that the propositional content of "Not every stick isn't under the rubber tree" is equivalent to that of utterance (2), but because it does not mention any individual stick, it does not allow whoever speaks next to make any reference to the stick that is under the gumgum tree.**

There are two characteristics of these dimensions and the multifaceted nature of utterances that introduce complications into natural language processing. First, as Halliday (1977) has pointed out, the units in which the information is conveyed along these other dimensions of meaning do not follow the constituent structure of sentences

*These dimensions parallel the three functions of language—ideational, interpersonal, and textual—in Halliday (1970), but the perspective I take on them is closer to that presented in Levy (1978).

*Pittenger et al. (1960) point out that "no matter what else human beings may be communicating about, or may think they are communicating about, *they are always communicating about themselves, about one another, and about the immediate context of the communication.*"

**This example is based on one suggested by Barbara Partee for the Sloan Workshop at the University of Massachusetts, December, 1978. A discussion of her example is included in Grosz and Hendrix, 1978.

nearly so nicely as do the units conveying propositional content. In particular, the social implications of an utterance are typically reflected in choices scattered throughout it; for example, they are reflected in the choice of utterance type (a request vs. a command) and in the choice of lexical items.

Second, an utterance may relate to plans and goals along any number of these dimensions. It may be a comment on the preceding utterance itself, its social implications (or both, as is usually the case with "I shouldn't have said that"), or on some part of the domain content of the utterance. It is not simply a matter of determining where an utterance fits into a speaker's plan, but of determining which plan or plans—domain, social, or discourse—the utterance fits into. A one dimensional analysis of an utterance is insufficient to capture the different effects (cf. Goffman, 1978).

The multifaceted nature of utterances poses problems for language generation as well. A speaker typically must coordinate goals along each of these dimensions. He must design an utterance that conveys information linking it to the preceding discourse and maintains the social relationship he has with the hearer(s) (or establishes one) as well as conveying domain-specific information.* The speaker's task is further complicated because he has only incomplete knowledge of the intended hearer's goals, plans and beliefs.

V. State of Art

I will use our work in natural language processing at SRI International (Robinson, 1980a; Walker, 1978) as an exemplar for discussing the current state of research in this area, both because I am most familiar with it and because I think the framework it provides is a useful one for seeing not only where the field stands, but also where the next several years effort might best be expended. A caveat is necessary before proceeding. The discussion that follows considers only research concerned with developing theoretical models of language use and the systems that contribute to this research. Because of space limitations, I will not discuss a second major direction of current research in natural language processing, that concerned with the construction of practical natural language interfaces (e.g., Hendrix, et al., 1978). The major difference between the two kinds of efforts is that research on interfaces has separated language processing from the rest of the system whereas one of the major concerns of research in the more theoretical direction is the interaction between language-specific and general knowledge and reasoning in the context of communication.

SRI's TDUS system has been contructed as part of a research effort directed at investigating the knowledge and processes needed for participation in task-oriented dialogues (Robinson, 1980a). The system participates in a dialogue with a user about the performance of an assembly task. It coordinates multiple sources of language-specific knowledge and combines them with certain general knowledge and common-sense reasoning strategies in arriving at a literal interpretation of an utterance in the context of an ongoing task-oriented dialogue.* A major feature of the system is the tight coupling of syntactic form and semantic interpretation. In the interpretation of an utterance, it associates collections of attributes with each phrase. For example, noun phrases are annotated with values for the attribute "definiteness," a property that is relevant for drawing inferences about focusing (Grosz, 1977a, 1977b, 1980) and about presuppositions of existence and mutual knowledge (Clark and Marshall, 1980).

Interpretation is performed in multiple stages under control of an executive and in accordance with the specifications of a language definition that coordinates multiple "knowledge sources" for interpreting each phrase. Two sorts of processes take part in the linguistic level of analysis. First, there are processes that interpret the input "bottom up" (i.e., words ⇒ phrases ⇒ larger phrases ⇒ sentences). In the analysis of utterance (2), these processes would provide attributes specifying that the phrase "a stick" is indefinite and in the subject position of a there-initial sentence. They would specify that the phrase "the rubber tree" is definite and presupposes the existence of a uniquely identifiable entity. Second, there are processes that refine the interpretation of a phrase in the context of the larger phrases that contain it, doing such things as establishing a relationship between syntactic units and *descriptions* of (sets of propositions about) objects in the domain model. For example, the structure for "the rubber tree" would include formal logical expressions regarding existence and treeness.

The assimilation level in the current system only goes so far as determining a literal interpretation in context. The major tasks performed here include delimiting the scope of quantifiers and associating references to objects with particular entities in the domain model, taking into account the overall dialogue and task context. To perform these tasks, the system includes mechanisms for representing and reasoning about complex processes (Appelt et al, 1980). In the case of our two monkeys, the system would determine whether there was a unique rubber tree in, or near, the focus of attention of the monkey (more on this shortly) and then posit, or check, the existence of a stick under it.

Although it only interprets utterances literally, TDUS does make some inferences based on the information explicitly

*Levy discusses how the multiple levels along which a speaker plans are reflected in what he says and the structure of his discourse.

*Several other systems are capable of fairly sophisticated analysis and processing at the level of coordinating different kinds of language-specific capabilities (e.g., Sager and Grishman, 1975; Landsbergen, 1976; Plath, 1976; Woods et al, 1976; Bobrow et al, 1977; Reddy et al, 1977) and of taking into account some of the ways in which context affects meaning through the application of limited action scenarios (Schank et al, 1975; Novak, 1977) or by considering (either independently or in conjunction with such scenarios) language-specific mechanisms that reference context (Hobbs, 1976; Rieger, 1975; Hayes, 1978; Mann et al, 1977; Sidner, 1979).

contained in an utterance. The plans it knows about are partially ordered (and not linear), and the structures it uses allow for describing plans at multiple levels of abstraction. To see the sorts of inferences TDUS will make, consider the sequence:

(6) User: I am attaching the pump.
(7) System: OK
(8) User: Which wrench should I use to bolt it?

In interpreting utterance (6), the sytem updates it's model of the task of attaching the pump. It uses tense and aspect information to determine that the task has been started but not completed (the user said, "am attaching," not "have attached."). As part of interpreting this utterance, the system also records that the user is now focusing on the pump and the attaching operation. The system uses this focusing information and information in its model of the task to determine that the bolting operation referred to is a substep of the attaching operation and that the "it" in utterance (8) is being used to refer to the pump. In addition, TDUS infers that all of the substeps of the attaching operation that had to precede the bolting have been done (Appelt, et al, 1980; Robinson, 1980b).

Initial progress has been made in overcoming the limitations of literal interpretation and including a consideration of a speaker's plans and goals in the interpretation of an utterance. Recent research on the role of planning in language processing includes that of Cohen (1978), Wilensky (1978), Carbonnell (1979), and Allen (1979). Cohen (1978) views speech acts (Searle, 1969) as one kind of goal-oriented activity and describes a system that uses mechanisms previously used for planning nonlinguistic actions to plan individual speech acts (on the level of requesting and informing) intended to satisfy some goals involving the speaker's or hearer's knowledge. In Wilensky's work on story understanding (see also Schank and Abelson, 1977), the speaker's overall plans and goals, some of which are implicit, are inferred from substeps and intermediate or triggering states (e.g., inferring from "John was hungry. He got in his car." that John was going to get something to eat.). Carbonell (1979) describes a system constructed to investigate how two agents with different goals interpret an input differently; it is particularly concerned with the effect of conflicting plans on interpretation. Allen (1979) describes a system based on a model in which speech acts are defined in terms of "the plan the hearer believes the speaker intended him to recognize" and has perhaps gone furthest in determining mechanisms by which a speaker's goals and plans can be taken into account in the interpretation of an utterance.

These efforts have demonstrated the feasibility of incorporating planning and plan recognition into the common-sense reasoning component of a natural language processing system, but their limitations highlight the need for more robust capabilities in order to achieve the integration of language-specific and general common-sense reasoning capabilities required for fluent communication in natural language. No system combines a consideration of multiple agents having different goals with a consideration of the problems that arise from mutiple agents having separate beliefs and each having only incomplete knowledge about the other agent's plans and goals.* Furthermore, only simple sequences of actions have been considered, and no attempt has been made to treat hypothetical worlds.

One of the major weaknesses in current AI systems and theories (and the limitation of current systems that I find of most concern) is that they consider utterances as having a single meaning or effect. Analogously, a critical omission in work on planning and language is that it fails to consider the multiple dimensions on which an utterance can have effects. If utterances are considered operators (where "operator" is meant in the general sense of something that produces an effect), they must be viewed as *conglomerate* operators.

Although it does not yet go beyond literal interpretation (except by filling in unmentioned intermediate steps in the task being performed), TDUS does account for two kinds of effects of an utterance. In addition to determining the propositional content of an utterance (and what it literally conveys about the state of the world), the system determines whether the utterance indicates that the speaker's focus of attention has shifted (Grosz, 1977a,b, 1980; Sidner, 1979).**

To summarize then, one or more of the following crucial limitations is evident in every natural language processing system constructed to date (although most of these problems have been addressed to some extent in the research described above and elsewhere):

- Interpretation is literal (only propositional content is determined).
- The knowledge and beliefs of all participants in a discourse are assumed to be identical.
- The plans and goals of all participants are considered to be identical.
- The multifaceted nature of utterances is not considered.

To move beyond this state, the major problems to be faced at the level of linguistic analysis concern determining how different linguistic constructions are used to convey information about such things as the speaker's (implicit) assumptions about the hearer's beliefs, what entities the speaker is focusing on, and the speaker's attitude toward the hearer. The problems to be faced at the assimilation level are more fundamental. In particular, we need to determine common-sense reasoning mechanisms that can derive complex connections between plans and goals—connections that are not explicit either in the dialogue or in the plans and

*Moore (1979) discusses problems of reasoning about knowledge and belief.

**Grosz and Hendrix (1978) discuss focusing as one of the elements of cognitive state crucial to the interpretation of both definite and indefinite referring expressions, and Grosz (1980) discusses several open problems in modeling the focusing process.

goals themselves—and to reason about these relationships in an environment where the problem solver's knowledge is necessarily incomplete. This is not just a matter of specifying more details of particular relationships, but of specifying new kinds of problem solving and reasoning structures and procedures that operate in the kind of environment in which natural language communication usually occurs.

VI. Common-Sense Reasoning in Natural Language Processing

The previous sections of this paper have suggested several complexities in the common-sense reasoning needs of natural language communication. A participant in a communicative situation typically has incomplete information about other participants. In particular he cannot assume that their beliefs, goals, or plans are identical. Communication is inherently interpersonal. Furthermore, the information a speaker conveys typically requires a sequence of utterances. As a result, interpretation requires recognition of different kinds of plans, and generation requires the ability to coordinate multiple kinds of actions to satisfy goals along multiple dimensions. Other complications are introduced by the interactions among plans of different agents (Bruce and Newman, 1978; Hobbs and Robinson (1978) discuss some of the complexity of the relationship between an utterance and domain specific plans).

From this perspective, the current deduction and planning systems in AI are deficient in several areas critical for natural language processing. A review of the current state of the art in plan generation and recognition shows that the most advanced systems have one or another (but not both) of the following capabilities: plans for partially ordered sequences of actions can be generated (Sacerdoti, 1977) and recognized (Genesereth, 1978; Schmidt and Sridhara, 1977) at multiple levels of detail in a restricted subject area. However, these programs only consider single agents, assume the system's view of the world is "the correct" one, and plan for actions that produce a state change characterized by a single primary effect.

The most important directions in which these capabilities must be extended and integrated for use in the interpretation and generation of language are the following:

- It must be possible to plan in a dynamic environment that includes other active agents, given incomplete information.
- It must be possible to coordinate different types of actions and plan to achieve mutiple primary effects simultaneously.
- It must be possible to recognize previously unanticipated plans.

VII. Conclusions

Common-sense reasoning, especially planning, is a central issue in language research, not only within artificial intelligence, but also in linguistics (e.g., Chafe, 1978; Morgan 1978), sociolinguistics (e.g., Kasher, 1978). The literal content of an utterance must be interpreted within the context of the beliefs, goals, and plans of the dialogue participants, so that a hearer can move beyond literal content to the intentions that lie behind the utterance. Furthermore, it is insufficient to consider an utterance as being addressed to a single purpose. Typically, an utterance serves multiple purposes: it highlights certain objects and relationships, conveys an attitude toward them, and provides links to previous utterances in addition to communicating some propositional content.

Progress toward understanding the relationship between utterances and objectives and its effect on natural language communication will be best furthered by consideration of the fundamental linguistic, common-sense reasoning, and planning processes involved in language use and their interaction. A merger of research in common-sense reasoning and language processing is an important goal both for developing a computational theory of the communicative use of language and for constructing computer-based natural language processing systems. The next few years of research on language processing should be concerned to a large extent with issues that are at least as much issues of common-sense reasoning (especially planning issues). While common-sense reasoning research could continue without any regard for language, there is some evidence that the perspective of language processing will provide insights into fundamental issues in planning that confront AI more generally.

Finally, I want to emphasize the long-term nature of the problems that confront natural language processing research in AI. I believe we should start by adding communication capabilities to systems that have solid capabilities in solving some problem (constructing such systems first if necessary; cf. McDermott, 1976). Although it may initially take longer to create functioning systems, the systems that result will be useful, not toys. People will have a reason to communicate with such systems. Monkey2 can help monkey1 get something to eat only if he himself has a realistic conception of the complexities of monkey1's world.

Acknowledgements

The Natural Language Research Group of the Artificial Intelligence Center at SRI International provides a stimulating environment in which to pursue research in language processing. Many of the ideas presented in this paper are the product of thought-provoking discussions, both oral and written, with members of this group, especially Gary Hendrix, Douglas Appelt, Jerry Hobbs, Robert Moore, Nils Nilsson, Ann Robinson, Jane Robinson, Earl Sacerdoti, and Donald Walker. I would also like to thank David Levy and Mitch Marcus for many discussions about language and the insights they provided. Many of these people also provided helpful comments on this paper.

References

(1) Allen, J. 1979. A Plan-Based Approach to Speech Act Recognition. Technical Report No. 131/79, Department of Computer Science, University of Toronto.

(2) Appelt, D. E., B. J. Grosz, G. G. Hendrix, and A. E. Robinson, 1980. "The representation and use of process knowledge." Technical Note 207. Artificial Intelligence Center, SRI International, Menlo Park, California.

(3) Bobrow, D. G., et al. 1977. GUS, A Frame Driven Dialog System. *Artificial Intelligence*, 8, 155-173.

(4) Bruce, B. and D. Newman. 1978. Interacting Plans. *Cognitive Science*, 2, 195-233.

(5) Carbonell, J. G. 1979. Computer Models of Social and Political Reasoning. Ph. D. Thesis, Yale University, New Haven, CT.

(6) Chafe, W. 1978. The Flow of Thought and the Flow of Language. In T. Givon (ed.), *Syntax and Semantics*, Vol. 12. Academic Press, New York.

(7) Clark, H. and C. Marshall. 1980. Definite Reference and Mutual Knowledge. In *Elements of Discourse Understanding: Proc. of a Workshop on Computational Aspects of Linguistic Structure and Discourse Setting*, A. K. Joshi, I. A. Sag,, and B. L. Webber, eds. Cambridge University Press, Cambridge, England.

(8) Cohen, P. 1978. On Knowing What to Say: Planning Speech Acts. Technical Report No. 118, Department of Computer Science, University of Toronto, Toronto, Canada.

(9) Collins, A. 1978. Studies of Plausible Reasoning. Volume I. BBN Report No. 3810, Bolt Beranek and Newman, Cambridge, MA.

(10) Donnellan, K. S. 1977. "Reference and Definite Descriptions." In S. P. Schwartz (Ed.), *Naming, Necessity, and Natural Kind*. Ithaca, New York: Cornell University Press. Pp. 42-65.

(11) Genesereth, M. R. 1978. Automated Consultation for Complex Computer Systems. Ph. D. Thesis, Harvard University, Cambridge, MA.

(12) Goffman, E. 1974. *Frame analysis*. New York: Harper.

(13) Goffman, E. 1978. Response Cries. *Language*, Vol. 54, 787-815.

(14) Grosz, B. 1977a. The Representation and Use of Focus in Dialogue Understanding. Technical Note 151, Artificial Intelligence Center, SRI International, Menlo Park, CA.

(15) Grosz, B. J. 1977b. The Representation and Use of Focus in a System for Understanding Dialogs. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*. Carnegie-Mellon University, Pittsburgh, PA. Pp. 67-76.

(16) Grosz, Barbara J. 1980. Focusing and description in natural language dialogues. In *Elements of Discourse Understanding: Proc. of a Workshop on Computational Aspects of Linguistic Structure and Discourse Setting*, A. K. Joshi, I. A. Sag, and B. L. Webber, eds. Cambridge University Press, Cambridge, England.

(17) Grosz, B. and G. Hendrix. 1978. A Computational Perspective on Indefinite Reference. Presented at Sloan Workshop on Indefinite Reference, University of Massachusetts, Amherst, MA, December 1978. Also Technical Note No. 181, Artificial Intelligence Center, SRI International, Menlo Park, CA (in preparation).

(18) Halliday, M.A.K. 1970. Language Structure and Language Function. In *New Horizons in Linguistics*, J. Lyons (ed.) (Penguin).

(19) Halliday, M.A.K. 1977. Structure and Function in Language. Presented at Symposium on Discourse and Syntax, University of California at Los Angeles, Los Angeles, CA, November 1977.

(20) Hayes, P. J. 1978. Some Association-Based Techniques for Lexical Disambiguation by Machine. Doctoral Thesis, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland.

(21) Hendrix, G. G., E.D. Sacerdoti, D. Sagalowicz, and J. Slocum. 1978. "Developing a Natural Language Interface to Complex Data." *ACM Transactions on Database Systems*, Vol. 3, no. 2, June 1978, pp. 105-147.

(22) Hobbs, J. 1976. A Computational Approach to Discourse Analysis. Research Report 76-2, Department of Computer Sciences, City College, City University of New York, New York, New York.

(23) Hobbs, J. 1979. Conversation as Planned Behavior. In *Proceedings of the Sixth International Joint Conference on Artificial Intelligence*. Stanford University, Stanford, CA.

(24) Hobbs, J. and J. Robinson 1978. Why Ask? SRI Technical Note 169, SRI International, Menlo Park, CA.

(25) Kasher, A. 1978. Indefinite Reference: Indispensability of Pragmatics. Presented at Sloan Workshop on Indefinite Reference, University of Massachusetts, Amherst, MA, December 1978.

(26) Labov, W. and D. Fanshel. 1977. *Therapeutic Discourse*. New York: Academic Press.

(27) Landsbergen, S. P. J., 1976. Syntax and Formal Semantics of English in PHLIQAI. In *Coling 76, Preprints of the 6th International Conference on Computational Linguistics.* Ottawa, Ontario, Canada, June 28-July 2, 1976. No. 21.

(28) Levy, D. 1978. Communicative goals and strategies: Between discourse and syntax. In T. Givon (ed.), *Syntax and Semantics*, Vol. 12. Academic Press, New York.

(29) Mann, W., J. Moore, and J. Levin 1977. A comprehension model for human dialogue. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, Carnegie-Mellon University, Pittsburgh, PA. Pp. 77-87.

(30) McCarthy, J. 1968. Programs with Common Sense. In M. Minsky (ed.) *Semantic Information Processing*, MIT Press, Cambridge, MA. Pp. 403-419.

(31) McDermott, D. 1976. Artificial Intelligence Meets Natural Stupidity. *SIGART Newsletter*, no. 57, Pp. 4-9.

(32) Moore, R. 1979. Reasoning about Knowledge and Actions. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

(33) Morgan, J. L. 1978. Two Types of Convention in Indirect Speech Acts. In Peter Cole (ed.), *Syntax and Semantics*, Vol. 9, Academic Press, Inc., New York, NY.

(34) Nilsson, N. 1971. *Problem Solving Methods in Artificial Intelligence*. McGraw-Hill, New York.

(35) Novak, G. 1977. Representations of knowledge in a program for solving physics problems. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence.* Carnegie-Mellon University, Pittsburgh, PA. Pp. 286-291.

(36) Pittenger, R., Hockett, C., and Danehy J., 1960. *The First Five Minutes*. Paul Martineau, Ithaca, NY.

(37) Plath, W. J. 1976. Request: A Natural Language Question-Answering System. *IBM Journal of Research and Development*, 20, 4, 326-335.

(38) Quine, W. V. O. 1960. *Word and Object*. MIT Press, Cambridge, MA.

(39) Reddy, D. R., et al. 1977. Speech Understanding Systems: A summary of Results of the Five-Year Research Effort. Department of Computer Science. Carnegie-Mell University, Pittsburgh, PA.

(40) Rieger, C. 1975. Conceptual Overlays: A Mechanism for the Interpretation of Sentence Meaning in Context. Technical Report TR-354. Computer Science Department, University of Maryland, College Park, MD.

(41) Robinson, A. 1980a. Understanding Natural-Language Utterances in Dialogs about Tasks. Technical Note 210. SRI International, Menlo Park, CA.

(42) Robinson, A. 1980b. Interpreting Verb Phrase References in Dialogs. To appear in *Proceedings of the Third Conference of the Canadian Society for Computational Studies of Intelligence*. Victoria, British Columbia. May 14-16, 1980.

(43) Rubin, A. D. 1978. A Theoretical Taxonomy of the Differences Between Oral and Written Language. In R. Spiro et al. (eds.) *Theoretical Issues in Reading Comprehension*. Lawrence Erlbaum, Hillsdale, NJ.

(44) Sacerdoti, E. D. 1977. *A Structure for Plans and Behavior*. Elsevier, NY.

(45) Sacerdoti, E. D. 1978. What Language Understanding Research Suggests about Distributed Artificial Intelligence. Proceedings of a workshop held at Carnegie-Mellon University, December 7-8, 1978.

(46) Sager, N. and R. Grishman. 1975. The Restriction Language for Computer Grammars. *Communicotions of the ACM*, 18, 390-400.

(47) Schank, R. C., and Yale A.I. Project 1975. SAM—A story understander. Research Report No. 43, Department of Computer Science, Yale University, New Haven, CT.

(48) Schank, R. and R. Abelson. 1977. *Scripts, Plans, Goals, and Understanding*. Laurence Erlbaum Associates, Hillsdale, NJ.

(49) Schmidt, C. F. and N. S. Sridharan. 1977. Plan Recognition Using a Hypothesis and Revise Paradigm: and Example. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, Carnegie-Mellon University, Pittsburgh, PA. Pp. 480-486.

(50) Searle, J. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, England.

(51) Sidner, C. L. 1979. A Computational Model of Co-Reference Comprehension in English. Ph. D. Thesis, Massachusetts Institute of Technolgy, Cambridge, MA.

(52) Walker, D. E. (ed.) 1978. *Understanding Spoken Language*. Elsevier North-Holland, NY.

(53) Wilensky, R. 1978. Understanding Goal-Based Stories. Ph.D. Thesis. Yale University, New Haven, CT.

(54) Woods, W. A., et al. 1976. Speech Understanding Systems: Final Report. BBN Report No. 3438, Bolt Beranek and Newman, Cambridge, MA.